

SPECIFICATION

TITLE OF THE INVENTION

METHOD OF AND COMPUTER SYSTEM FOR EXECUTING AREA-DIVISION AND COMPRESSION OF DOCUMENT IMAGE

5

[0001] This application is based on No. 2003-94230 filed in Japan, the contents of which are hereby incorporated by reference.

10 BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0002] The present invention relates to a method of and a computer system for executing area-division and compression of a document image.

15 DESCRIPTION OF PRIOR ART

[0003] Generally, in an image processing system for digital printing, a document image is read using a scanner or the like to temporarily store image data corresponding to the document image onto a memory, and then an image is printed based on the image data stored on the memory. In order to efficiently utilize a memory limited in storage capacity, or reduce the storage capacity of a memory to be installed in the image processing system, or increase the transmission rate of the image data, the image processing system is configured such that the image data is stored in

20
25

a compressed form, and the stored image data is expanded in the image-printing process.

[0004] As for the document image, there are various types, such as photograph and graphic, different in attribute (image characteristic), and a compression technique optimal to each image data of such document images is varied depending on their attributes. Therefore, in this type of conventional image processing system, a plurality of compression techniques are typically employed, and selectively used in conformity to each of the attributes of the document images.

[0005] Further, a document image on a single sheet can include a plurality of images different in attribute in a mixed manner. In such a case, if the plural types of image data of the document image are simply compressed using a single compression technique, the entire document image cannot be optimally compressed. From this point of view, there has been proposed an image compression technique in which a document image is divided into a plurality of areas based on their respective attributes, and each of the areas is individually compressed in conformity to the attribute thereof (see, for example, Japanese Laid-Open Patent Publications No. 2001-136521 and No. 10-108011).

[0006] According to the above conventional compression technique, only if a plurality of areas different in

attribute are separated or independent from each other in a document image on a single sheet, respective image data of the areas can be optimally compressed. However, as the result of a process of dividing a document image into a plurality of areas based on their respective attributes, one area associated with a first attribute is likely to partly or fully overlap with another area associated with a second attribute different from the first attribute. In other words, a certain area has a plurality of attributes, and all of the plural attributes are overlappingly recognized for this area in the area division process. In such a case, the above conventional image-data compression technique cannot identify the attribute of the overlap area, and consequently cannot determine any specific compression technique to be used for the overlap area.

SUMMARY OF THE INVENTION

[0007] In view of the above conventional problem, it is therefore an object of the present invention to provide an image-data compression technique for use in a process of dividing a document image with a plurality of mixed images different in attribute, into a plurality of areas, capable of performing optimal compression even in an overlap area where the areas partly or fully overlap with each other.

[0008] According to an aspect of the present invention,

there is provided a method of executing area-division/compression of a document image, which may be automatically executed by a computer using a program for executing the method. The method includes first to fourth steps as described below.

[0009] In the first step, a document image is sectionalized (or divided) on the basis of attributes associated with a compression mode to form a plurality of image areas each having an attribute. In the second step, it is determined whether the image areas partly or fully overlap with one another to form an overlap area. In the third step, if the overlap area is formed, the overlap area is separated from the image areas overlapping with one another (hereinafter referred to as "engaged areas"), and then an attribute associated with the compression mode for the overlap area is newly set according to priorities of the respective attributes of the engaged areas, and the positional relationship between the overlap area and each of the engaged areas. In the fourth step, each of the image areas and the overlap area is compressed using a compression technique corresponding to each of their attributes.

[0010] In the third step, if one of the engaged areas is included within the remaining image areas, the attribute for the overlap area may be set up according to an

inclusive relationship of the engaged areas. Alternatively, the attribute for the overlap area may be set up according to the occupancy rate of the overlap area to each of the engaged areas, and weighting coefficients of the respective attributes of the engaged areas. Further, in the fourth step, the overlap area may be compressed by a combination of the compression technique for the respective attributes of the engaged areas and the compression technique for the attribute of the overlap area.

[0011] According to another aspect of the present invention, there is provided a computer system for executing area-division/compression of a document image. The computer system includes first to fourth sections which perform the first to fourth steps of the above-mentioned method, respectively.

[0012] According to the method or computer system of the present invention, an optimal area division can be executed while minimizing erroneous determinations. In addition, any overlap area can be eliminated to avoid any unproductive processing in compression/storage processing while allowing an optimal compression method to be applied thereto in consideration of the respective attributes (characteristic) of the divided areas. Furthermore, the need for acquiring information about surroundings of the overlap area can be eliminated to achieve a high-speed

image data processing.

[0013] According to the present invention, in a process of dividing a document image including a plurality of mixed images different in attribute, into a plurality of areas, and compressing image data of the areas, an optimal compression can be performed even in an overlap area where the areas partly or fully overlap with one another.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] Other features and advantages of the present invention will be apparent from the accompanying drawings and from the detailed description.

[0015] FIG. 1 is a block diagram showing one example of an image processing systems for executing an area-division/compression processing according to one embodiment of the present invention.

[0016] FIG. 2 is a flowchart showing a procedure of area-division/compression according to one embodiment of the present invention.

[0017] FIG. 3 is a flowchart showing a procedure of area-division/compression according to one embodiment of the present invention.

[0018] FIG. 4 is a flowchart showing a procedure of area-division/compression according to one embodiment of the present invention.

[0019] FIG. 5 is an explanatory view of one example of an area division technique for use in case where two areas overlap with one another.

5 [0020] FIG. 6 is an explanatory view of one example of an area division technique for use in case where three areas overlap with each other.

[0021] FIG. 7 is an explanatory view of another example of the area division technique for use in case where three areas overlap with each other.

10 [0022] FIG. 8 is an explanatory view of one example of an area division technique for use in case where a photograph area partly overlaps with a letter area.

[0023] FIG. 9 is an explanatory view of one example of a printed image in case where an overlap area between a photograph area and a letter area is defined as a letter area.
15

[0024] FIG. 10 is an explanatory view of another example of the area division technique for use in case where a photograph area partly overlaps with a letter area.

20 [0025] FIG. 11 is an explanatory view of one example of a printed image in case where an overlap area between a photograph area and a letter area is defined as a photograph area.

25 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0026] With reference to the drawings, an embodiment of the present invention will now be specifically described.

[0027] FIG. 1 is a block diagram showing one example of an image processing system for executing a document-image area-division/compression processing according to the present invention. As shown in FIG. 1, this image processing system includes an input section 5 for entering image data therethrough, a personal computer 1 capable of executing various programs, and an output section 8 for outputting processed image data. In the image processing system, all of the personal computer 1, a scanner 7 capable of reading a document image at a desired resolution, a server capable of storing various data, such as image data and program data, to serve as an external storage device, and a printer 10 for printing out on a recording sheet are connected with each other via a local area network n (LAN).

[0028] The personal computer 1 is equipped with a monitor 3 for displaying data, and an input device 4 including a keyboard, a mouse and others. The personal computer 1 also incorporates therein a hard disk 2 as a built-in storage device, and a Data Base is established in the personal computer 1 including the hard disk 2. This Data Base includes various databases, such as an area-information database, an area-attribute changing table, and a compression database. Further, the personal computer 1

includes a processing module for performing an area-division/compression processing according to the present invention, and another processing module for performing other conventional image processing.

5 [0029] While this embodiment will be described in connection with a system in which all of an input device, an image processing device and an output device are connected with each other via a local area network n, the present invention may be executed using an apparatus
10 constructed by integrally combining a scanner (input section), an image processing section and a printer (output section), such as an MFP (Multiple Function Peripheral).

[0030] In the image processing system, image data corresponding to a document image is entered from the input
15 section 5 into the personal computer 1. The input section 5 includes the server 6 for supplying image data, and the scanner 7 for reading a document image and converting the read document image into image data. The personal computer 1 is operable to output image data stored in the Data Base
20 (hard disk 2) to the output section 8. The output section 8 includes the printer 10 for printing an image according to the output image data.

[0031] A document-image area-division/compression technique using the personal computer 1 will be described
25 below. First of all, an outline or fundamental concept of

the document-image area-division/compression will be described. The personal computer 1 executes the following procedure according to an area-division/compression program to perform the document image area-division/compression.

5 Specifically, the personal computer 1 first sectionalizes (or divides) a document image into a plurality of image areas each having an attribute associated with a compression mode. Then, the personal computer 1 determines whether the image areas partly or fully overlap with each
10 other to form an overlap area.

[0032] When the formation of an overlap area is determined, the overlap area is separated (or divided) from the engaged areas or the image areas overlapping with each other. Then, an attribute associated with one compression
15 mode is newly set up in the overlap area according to priorities of the respective attributes of the engaged areas, and the positional relationship between the overlap area and each of the engaged areas. For example, if one image area in the engaged areas is included within the
20 remaining image area(s), an attribute will be set up in the overlap area according to the inclusive relationship of the engaged areas. Otherwise, if there is not such an inclusive relationship, an attribute will be set up in the overlap area according to the occupancy rate of the overlap
25 area to each of the engaged areas, and weighting

coefficients of the respective attributes of the engaged areas. After the setting of the attribute of the overlap area, each of the image areas and the overlap area is compressed using a compression technique corresponding to each of their attributes.

[0033] With reference to the flowcharts in FIGS. 2 to 4, a specific processing routine of the document-image area-division/compression processing to be executed by the personal computer 1 will be described below. As shown in FIGS. 2 to 4, in this processing routine, a document image is first read from the input section 5 (server 6, scanner 7) at Step S1.

[0034] Subsequently, at Steps S2 to S4, an area-determination/division processing is executed with respect to photograph, letter and graphic. That is, this processing routine is configured such that the document image is sectionalized (or divided) into a photograph area, a letter area and a graphic area, which are different in optimal compression technique. In this embodiment, the compression techniques for used in a normal case where the image areas are separated independent from each other are shown in Table 1.

[0035]

Table 1 Compression Technique in Normal State

	Letter Area	Graphic Area	Photograph Area
Compression Technique	MMR	Flate	Jpeg (regular)

[0036] Specifically, the letter area is compressed using
5 an MMR adapted to perform data compression through
binarization to provide the highest data compression ratio.
The graphic area is compressed using a Flate adapted to
perform reversible data compression. The photograph area
is compressed using a regular Jpeg adapted to perform
10 nonreversible data compression.

[0037] Then, at Step S5, the area information obtained
through the area-determination and area-division steps is
stored in the area-information database in the Data Base
(hard disk 2). Further, based on the area information
15 stored in the area-information database, it is checked
whether there is any overlap area. When the presence of
the overlap area is determined, information about the
overlap area is stored in the area-information database.

[0038] Then, at Step S7, a leading area in the three
20 areas is retrieved (initialization). Subsequently, at Step
S8, based on the area information stored in the area-
information database, the leading area is extracted. After
the area is extracted, it is determined whether the

extracted area has a zone overlapping with either one of the remaining areas, at Step S9. When the area has an overlap zone (YES), positional information of the overlap zone or overlap area is acquired at Step S10. At Step S9, 5 if the area has no overlap zone (NO), the process will skip to Step S20.

[0039] Then, at Step S11, it is determined whether the overlap area is included within another one of the remaining areas. When the overlap area is not included 10 therewithin (NO), it is determined whether the two engaged areas causing the overlap area have the same attribute value (attribute), at Step S12. When the two engaged areas have the same attribute value (YES), Step S13 is executed to separate (or divide) this overlap area from the two 15 engaged areas. In this case, the separated overlap area has the same attribute value as that of the two engaged areas. Then, at Step S20, a record of the area-attribute change is registered (stored) in the area-information database.

20 [0040] At the above Step S11, if it is determined that the overlap area is included within another one of the remaining areas (YES), Steps S14 to S19 will be sequentially executed to set up respective attributes of the overlap area (inner area or innermost area) and the 25 area (outer area or outermost area) including the overlap

area in accordance with the inclusive relationship. More specifically, the attribute of the outer area is first retrieved at Step S14, and then the attribute of the inner area is retrieved at Step S15.

5 [0041] At Step S16, the area-attribute changing table in the Data Base (hard disc 2) is referred thereto. At Step S17, the outer area and the inner area are separated from one another. The attribute of the outer area and the attribute of the inner area are changed at Steps S18 and
10 S19, respectively. Then, at Step S20, a record of the area-attribute change is registered (stored) in the area-information database.

[0042] Table 2 shows a specific example of a technique for changing such attribute values.

15 [0043]

Table 2 Division/Attribute-Value Change of
Areas involved in Inclusive Relationship

		Case 1	Case 2	Case 3	Case 4
Before Change	Area 1 (including)	Photograph	Photograph	Photograph	Letter
	Area 2 (included)	Photograph	Letter	Graphic	Photograph
After Change	Area 1 (including)	Graphic	Photograph	Graphic	Letter
	Area 2 (included)	Photograph	Photograph	Graphic	Graphic

[0044] In Table 2, Case 1 is an example in which an area
20 2 associated with the photograph attribute is includes within an area 1 associated with the photograph attribute.

In this case, the attribute of the area 1 is changed to the graphic attribute because there is a high possibility that the attribute of the area 1 has been erroneously determined as the photograph attribute. Case 2 is an example in which an area 2 associated with the letter attribute is includes within an area 1 associated with the photograph attribute. In this case, the attribute of the area 2 is changed to the photograph attribute because the area 2 will be embossed if it is maintained in the letter attribute. Case 3 is an example in which an area 2 associated with the graphic attribute is includes within an area 1 associated with the photograph attribute. In this case, the attribute of the area 1 is changed to the graphic attribute because there is a high possibility that the attribute of the area 1 has been erroneously determined as the photograph attribute. Case 4 is an example in which an area 2 associated with the photograph attribute is includes within an area 1 associated with the letter attribute. In this case, the attribute of the area 2 is changed to the graphic attribute because there is a high possibility that the attribute of the area 2 has been erroneously determined as the photograph attribute.

[0045] In the case where the overlap area is not included within another one of the remaining areas (NO at Step S11), if the two engaged areas are different in

attribute value, Steps S21 to S27 will be sequentially executed to newly set an attribute value in the overlap area. More specifically, at Step S21, an areal rate Rate_(i) of the overlap area to each of the two engaged area is calculated by the following formula 1:

$$\text{Rate_}(i) = \text{NewAreaSize} / \text{AreaSize } (i)$$

, wherein NewAreaSize is the size of the overlap area, and AreaSize (i) is the size of an area having an area number (i).

10 [0046] At Step S22, the area-attribute changing table in the Data Base (hard disc 2) is referred thereto. At Step S23, an evaluation value Eval_(i) to the aerial rate Rate_(i) is generally calculated by the following formula 2:

$$15 \quad \text{Eval_}(i) = \text{func } (\text{Rate_}(i), \text{Coef_}(a))$$

, wherein (a) is an area attribute, Coef_(a) being a weighting coefficient set up with respect to each of the area attributes, and func (x, y) being an evaluation function.

20 [0047] Practically, the evaluation value Eval_(i) can be calculated, for example, by the following simple formula 3:

$$\text{Eval_}(i) = [(1 - \text{Rate_}(i)) * \text{Coef_}(a)]$$

[0048] Table 3 show a specific example of the weighting coefficient Coef_(a).

25 [0049]

Table 3 Weighting Coefficient of each Area (for
calculation of evaluation value)

Area	Weighting Coefficient
Photograph	1.5
Graphic	2
Letter	3

[0050] The above specific example is set up on the
 5 assumption that the processing is performed to a document
 image. Thus, the largest weighting coefficient is given to
 the letter area, and the smallest weighting coefficient is
 given to the graphic area. If it is set up on the
 assumption that the processing is performed to a general
 10 photograph, the largest and smallest weighting coefficients
 may be given to the photograph and letter areas,
 respectively.

[0051] That is, the above calculation method of the
 evaluation value $Eval_i$ to each of the areas is intended
 15 to calculate in which of the engaged areas the overlap area
 is included at a larger rate, so as to estimate to which of
 the attributes of the engaged area the attribute of the
 overlap area is close. For example, when only a small
 portion of the letter area is included in the photograph
 20 area, it is estimated that the overlap area has no
 attribute of photograph. This allows a printed image to
 give an impression in conformity to that to be made from

viewing of an original document image.

[0052] At Step S24, a maximum value of the evaluation value Eval_(i) is detected. At Step S25, an attribute value corresponding to the maximum value of the evaluation value Eval_(i) is detected. At Step S26, the overlap area is divided. At Step S27, an attribute value is newly given to the separated overlap area. Then, Step S20, a record of the area-attribute change is registered (stored) in the area-information database.

[0053] After the area-attribute change record is registered in the area-information database, it is determined whether any succeeding area remains. If any succeeding area remains (YES), Steps S8 to S28 will be repeatedly executed (a part of the steps may be not executed depending on conditions). When it is determined that no succeeding area remains (NO), the leading area is retrieved (initialization) at Step S29, and subsequently the leading area is extracted from the area-information database at Step S30.

[0054] Then, at Step S31, the area-information change record is acquired from the area-information database. Subsequently, at Step S32, a compression parameter is acquired from the compression database in the Data Base (hard disk 2). Then, at Step S33, the area is compressed according to the compression parameter acquired at Step S32.

[0055] Table 4 shows a specific example of a compression method for the area having the attribute change record.

[0056]

Table 4 Compression Method for Area having Attribute

5

Change Record

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
Attribute 1	Photo	Photo	Photo	Graphic	Graphic	Graphic	Letter	Letter	Letter
Attribute 2	Letter	Photo	Graphic	Letter	Photo	Graphic	Letter	Photo	Graphic
Compression Method	color reduction Flate	Jpeg (strong)	Jpeg (weak)	color reduction Flate	Jpeg (regular)	Flate	MMR	Jpeg (regular)	Flate

Attribute 1: initial area-attribute

Attribute 2: area attribute after re-area-division

[0057] Then, at Step S34, it is determined whether any succeeding area remains. If any succeeding area remains, Steps S30 to S34 will be repeatedly executed.

[0058] When it is determined whether no succeeding area remains at Step S34, an image data file is made from the image data compressed at Step S35. Then, at Step S36, the image data file is output to the output section 8 (server 6, printer 10), and the document-data area-division/compression processing is completed.

[0059] Some specific examples of the area division based on the area-division/compression processing as shown in FIGS. 2 to 4 will be described below.

[0060] As shown in FIG. 5, in case where two areas 1 and

2 partly overlap with one another but the overlap area is includes within neither the area 1 nor the area 2, when the areas 1, 2 are different in attribute value from one another, the overlap area is separated as an area 3, and an attribute value is newly set in the area 3, according to Steps S21 to S27. For example, if an evaluation value of the area 3 (overlap area) to the area 1 and an evaluation value of the area 3 to the area 2 are calculated by the formulas 1 to 3, and the former evaluation value is greater than the latter evaluation value, an attribute value identical to that of the area 1 will be given to the area 3.

[0061] Further, as shown in FIG. 6, in case where three areas 1, 2 and 3 partly overlap with each other but one (triple overlap area) of the overlap areas is includes within none of the areas 1, 2 and 3, when the areas 1, 2, 3 are different in attribute value from each other, this overlap area is separated as an area 4, and an attribute value is newly set in the area 4, according to Steps S21 to S27. For example, if an evaluation value of the area 4 (overlap area) to the area 1 and an evaluation value of the area 4 to the other areas 2, 3 are calculated by the formulas 1 to 3, and the former evaluation value is greater than the latter evaluation value, an attribute value identical to that of the area 1 will be given to the area 4.

[0062] Furthermore, when three areas (as well as four or

more areas) overlap with each other, and one of the areas is included within the overlap area between the remaining two areas, the overlap area between the two non-included area may be subjected to the area division processing to
5 determine an attribute value thereof, and then an attribute for each of this overlap area and the included area may be determined according to the inclusive relationship.

[0063] Specifically, as shown in FIG. 7, when three areas 1, 2 and 3 overlap with each other, and the area 3 is
10 included within the overlap area between the areas 1 and 2, the areas 1 and 2 are subjected to the same time division processing as that in FIG. 5. That is, when the areas 1, 2 are different in attribute value, the overlap area between the areas 1 and 2 is separated as an area 4 according to
15 Steps S21 to S27, and an attribute value is newly set up in the area 4. For example, if an evaluation value of the area 4 to the area 1 and an evaluation value of the area 3 to the other areas 2 are calculated by the formulas 1 to 3, and the former evaluation value is greater than the latter
20 evaluation value, an attribute value identical to that of the area 1 will be given to the area 4. Then, the attribute value is finally determined according to the inclusive relationship according to Steps S14 to S19. In the example illustrated in FIG. 7, the attribute value of
25 the area 4 is changed to the attribute value of the area 3.

[0064] A more specific area-division technique will be described below in conjunction with two cases where a single document image includes photograph and letter areas in a mixed manner, and the two areas partly overlap with one another.

[0065] (CASE 1)

As shown in FIG. 8, a photograph area having a vertical size of 100 and a horizontal size of 100 and a letter area having a vertical size of 20 and a horizontal size of 80 partly overlap with one another to form an overlap area having a vertical size of 20 and a horizontal size of 35. In this case, according to the formula 1, the areal rate of the overlap area to the photograph area Th_Photo is 0.07 ($(35 * 20) / (100 * 100) = 0.07$), and the areal rate of the overlap area to the letter area Th_Letter is 0.4375 ($(35 * 20) / (20 * 80) = 0.4375$).

[0066] Thus, according to the formula 3, the evaluation value of the overlap area to the photograph area The_Photo is 1.4895 ($(1 - 0.07) * 1.5 = 1.4895$), and the evaluation value of the overlap area to the letter area The_Letter is 1.6875 ($(1 - 0.4375) * 3 = 1.6875$). Therefore, the evaluation value The_Letter is greater than the evaluation value The_Photo , and thereby the overlap area is determined as a letter area.

[0067] FIG. 9 shows one example of a printed image using

an image data compressed under the condition that an overlap area is determined as a letter area in a document image including a photograph area and a letter area which partly overlap with one another to provide a greater evaluation value of the overlap area to the letter area than an evaluation value of the overlap area to the photograph area. As seen in FIG. 9, the printed image is adequately formed without any feeling of strangeness.

[0068] (CASE 2)

As shown in FIG. 10, a photograph area having a vertical size of 100 and a horizontal size of 100 and a letter area having a vertical size of 35 and a horizontal size of 120 partly overlap with one another to form an overlap area having a vertical size of 35 and a horizontal size of 85. In this case, according to the formula 1, the areal rate of the overlap area to the photograph area Th_Photo is 0.2975 $((85 * 35) / (100 * 100) = 0.2975)$, and the areal rate of the overlap area to the letter area Th_Letter is 0.7083 $((85 * 35) / (35 * 120) = 0.7083)$.

[0069] Thus, according to the formula 3, the evaluation value of the overlap area to the photograph area The_Photo is 1.50375 $((1 - 0.2975) * 1.5 = 1.50375)$, and the evaluation value of the overlap area to the letter area The_Letter is 0.8715 $((1 - 0.7083) * 3 = 0.8715)$.

Therefore, the evaluation value The_Photo is greater than

the evaluation value The_Letter, and thereby the overlap area is determined as a photograph area.

[0070] FIG. 11 shows one example of a printed image using an image data compressed under the condition that an overlap area is determined as a photograph area in a document image including a photograph area and a letter area which partly overlap with one another to provide a greater evaluation value of the overlap area to the photograph area than an evaluation value of the overlap area to the letter area. As seen in FIG. 11, the printed image is adequately formed without any feeling of strangeness.

[0071] The above image processing system may be configured such that the overlap area is compressed by a combination of the compression technique for the respective attributes of the image areas overlapping with each other and the compression technique for the attribute of the overlap area.

[0072] As mentioned above, according to the above embodiment of the present invention, an optimal area division can be executed while minimizing erroneous determinations. In addition, any overlap area can be eliminated to avoid any unproductive processing in compression/storage processing while allowing an optimal compression method to be applied thereto in consideration

of the respective attributes of the divided areas. Furthermore, the need for acquiring information about surroundings of the overlap area can be eliminated to achieve a high-speed image data processing.

- 5 [0073] An advantageous embodiment of the invention has been shown and described. It is obvious to those skilled in the art that various changes and modifications may be made therein without departing from the spirit and scope thereof as set forth in appended claims.